

# Réactualisation d'une image de référence pour la détection du mouvement dans les scènes urbaines

## Updating a Reference Image for Detecting Motion in Urban Scenes

par Patrick VANNOORENBERGHE\*, Cina MOTAMED\* et Jack-Gérard POSTAIRE\*\*

\*Laboratoire d'Analyse des Systèmes du Littoral,  
Université du Littoral-Côte d'Opale  
195, rue du Pasteur M.-L. King, BP 649, 62228 Calais CEDEX, France.  
Tél : 03-21-46-06-85, Fax : 03-21-46-06-86.  
e-mail : Patrick.Vannoorenberghe@asl-gw.univ-littoral.fr, motamed@univ-littoral.fr

\*\* Laboratoire d'Automatique  $I^3D$ ,  
Interaction, Image et Ingénierie de la Décision  
Université des Sciences et Technologies de Lille, Bâtiment P2  
59655 Villeneuve d'Ascq CEDEX, France.  
Tél : 03-20-43-42-88, Fax : 03-20-43-65-67. e-mail : jgp@cal.univ-lille1.fr

### *résumé et mots clés*

Nous présentons dans cet article une méthode de construction et de mise à jour d'une image de référence pour la détection du mouvement dans une scène urbaine. L'algorithme de détection proposé exploite les différences entre les contours statiques contenus dans la scène et les contours extraits de chaque image de la séquence. Il permet de mettre en évidence les contours des objets mobiles et les contours des zones affectées par le mouvement dans le cas où le fond n'est pas uniforme. L'image de référence, contrôlée de manière robuste, englobe les informations spatiales et temporelles contenues dans un grand nombre d'images successives de la séquence. Actualisée localement par l'intermédiaire d'un filtre récursif, elle permet d'intégrer, après un temps contrôlé, les objets qui s'arrêtent dans la scène. Ce type de réactualisation permet de s'adapter aux mouvements imprévisibles des objets (changement de vitesse et de direction) déformables ou non. De plus, l'analyse des contours a permis de s'affranchir des variations globales de l'éclairage ainsi que de l'hétérogénéité des fonds de la scène en milieu urbain. Cette approche obtient des résultats satisfaisants sur des images de scènes d'extérieur complexes.

Séquence d'images dynamiques, détection de mouvement, image de référence, filtre récursif, scènes d'extérieur.

### *abstract and key words*

We present in this paper a construction and updating method of a reference image for motion detection in an urban environment. The proposed detection algorithm exploits differences between static edges of the scene and edges extracted from each image of the sequence. It allows to detect moving objects contours and moving areas contours if the background is not uniform. The reference image, robustly controlled, contains spatio temporal information of a great number of successive images. Updated locally with a recursive filter, it allows to integrate, after a controlled time, objects that stop in the scene. This kind of updating permits to automatically adapt with unpredictable movements of deformable or rigid objects (changes in speed and direction). In addition, analyzing edges allows to take into account global illumination changes and heterogeneity of the scene background in an urban environment. This approach gives good results on complex outdoor image sequences.

Dynamic image sequence, motion detection, reference image, recursive filter, outdoor scenes.

## 1. introduction

La surveillance automatique de carrefours urbains est d'une grande importance pour une gestion optimale du trafic [11, 9]. Ces dernières années, les caméras de surveillance, initialement utilisées pour enregistrer les images de trafic, deviennent désormais les entrées potentielles de systèmes de vision artificielle et permettent l'analyse automatique du trafic urbain. Cependant, une segmentation directe de ces images est une tâche relativement complexe. La détection du mouvement a ainsi été suggérée comme méthode de segmentation [8, 6]. De plus, certaines études montrent que l'analyse des contours joue un rôle important dans les processus de surveillance automatique où la reconstruction du masque complet des objets en mouvement n'est pas nécessaire [17, 3].

Basés sur un travail original proposé par Haynes [18], plusieurs algorithmes de détection de mouvement ont été développés [5, 7]. Ces algorithmes prennent en compte deux ou trois images successives de la séquence en les combinant de manière à détecter les contours des objets en mouvement. Nous proposons un algorithme qui utilise un grand nombre d'images successives de la séquence de façon à détecter les mouvements imprévisibles des objets évoluant dans la scène. Nous nous plaçons ici dans un contexte où la caméra est fixe et les objets évoluent dans une scène où le fond est de nature complexe et non homogène. La méthode proposée est basée sur la comparaison entre une image contenant tous les contours statiques de la scène et chaque image de la séquence. Elle permet de mettre en évidence les zones affectées par le mouvement sous la forme des contours des zones affectées par le mouvement et des contours des objets mobiles. Le domaine applicatif englobe les systèmes de surveillance de carrefours urbains [16] en vue d'extraire des paramètres du trafic tels que des taux d'occupation ou l'analyse des flux, mais aussi la surveillance de sites sensibles pour le contrôle des accès (détection de présence) [15]. Notre contribution a consisté à développer un détecteur de mouvement capable de retrouver, dans chaque image de la séquence, les zones affectées par le mouvement en s'adaptant aux mouvements non uniformes des objets dans la scène.

La construction d'une image de référence est abordée dans la littérature scientifique par de nombreux auteurs [10, 1]. L'utilisation d'un filtre de Kalman permet de prendre en compte l'évolution de l'image au cours du temps [12]. Le réglage des gains du filtre est cependant délicat et ne permet pas de fixer le temps nécessaire pour l'intégration des objets dans l'image de référence. Enfin, plusieurs études montrent que l'utilisation d'une image de référence joue un rôle important dans les applications de surveillance et de détection d'intrusions [4, 2].

Cet article se décompose en quatre principales sections. Le principe de base de la méthode de détection de mouvement est présenté à la section 2. L'algorithme proposé exploite les changements entre une image de référence et chaque image de la séquence. La section 3 présente la construction et la réactualisation de cette image de référence. La décision concernant la détection d'un mouvement au niveau de chaque pixel de l'image résulte d'une procédure de seuillage local (section 4). Enfin, plusieurs exemples tirés de séquences urbaines réelles sont présentés et commentés à la section 5.

## 2. principe de base

L'algorithme de détection de mouvement proposé dans cet article est basé sur la comparaison de chaque image courante de la séquence avec une image de référence [14]. Compte-tenu des variations d'éclairage qui peuvent intervenir et des éventuelles modifications du fond de la scène, une procédure de réactualisation automatique de cette image est indispensable (cf. section 3). Soit  $I$  la séquence d'images originales, composée de  $N$  images, toutes définies sur une grille rectangulaire de dimensions  $X \times Y$ . Les valeurs des niveaux de gris des pixels de chaque image appartiennent à l'ensemble  $\{0, 1, \dots, N_g - 1\}$ . Nous désignons par  $I^k(P)$  le niveau de gris du pixel  $P$  dans l'image  $I^k$  de rang  $k$ . Un opérateur différentiel est appliqué sur chacune des  $N$  images de la séquence  $I$ . Le rôle de cet opérateur est de mettre en évidence les contours statiques contenus dans l'image. Notons  $G^k$  le gradient de l'image  $I^k$ , c'est-à-dire l'image constituée des valeurs de l'amplitude du gradient calculé en chaque point de  $I^k$ . On considère  $G^k$  comme l'**image courante des contours**. A chaque rang  $k$  de la séquence, l'image courante des contours  $G^k$  est comparée à l'image de référence de rang  $k$  notée  $R^k$ .  $R^k$  est l'image de tous les contours statiques de la scène, c'est-à-dire les contours du fond ainsi que les contours des objets statiques. On considère  $R^k$  comme l'**image de référence des contours**.

### 2.1. détection des points de contours affectés par le mouvement

Une **image des variations de gradient**, notée  $D^k$ , est construite pour chaque rang  $k$  de la séquence, de manière à mettre en évidence les différences entre l'image de référence des contours  $R^k$  et l'image courante des contours  $G^k$ . Pour construire cette image  $D^k$ , nous réalisons la différence pixel à pixel de  $R^k$  et de  $G^k$ .  $D^k(P)$ , la valeur de l'image  $D^k$  au pixel  $P$ , est calculée de la manière suivante :

$$D^k(P) = \frac{1}{A} \cdot [R^k(P) - G^k(P)]^2. \quad (1)$$

Dans cette expression,  $A$  désigne la constante de normalisation qui permet de recalculer les valeurs des différences élevées au carré dans l'ensemble  $\{0, 1, \dots, N_g - 1\}$ . Pour une séquence d'images codées sur 8 bits ( $N_g = 2^8$  niveaux), la constante  $A$  est égale à  $(2^8 - 1)$ . La valeur de  $D^k$  au pixel  $P$  est faible lorsqu'un contour apparaît à cet endroit simultanément dans l'image de référence des contours  $R^k$  et dans l'image courante des contours  $G^k$ .

### 2.2. extraction des points de contours affectés par le mouvement

De manière à séparer les pixels appartenant réellement à des contours affectés par le mouvement des pixels représentant des points du fond, l'image  $D^k$  est seuillée. A chaque pixel  $P$  est associé un seuil  $s^k(P)$  dont la valeur dépend de la stabilité

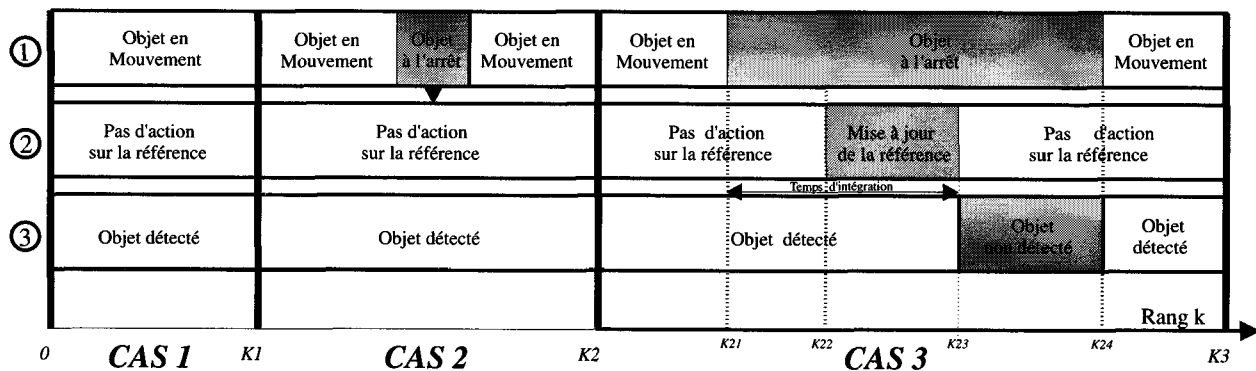


Figure 1. – Critères de réactualisation : 3 cas.

temporelle de l'image de référence des contours à cet endroit. Ce seuillage local adaptatif permet d'extraire l'ensemble des points de contours affectés par le mouvement des objets de chaque image de la séquence sous la forme d'une image binaire notée  $C^k$ . On considère  $C^k$  comme l'image des points de contours affectés par le mouvement. Pour chaque pixel  $P$ , l'image  $C^k$  est construite selon la règle de décision suivante :

$$C^k(P) = \begin{cases} 1 & \text{si } D^k(P) > s^k(P) \\ 0 & \text{sinon.} \end{cases} \quad (2)$$

On admet donc qu'un point de contour en mouvement est présent au pixel  $P$  dans l'image  $C^k$  si la valeur de l'image des variations de gradient à cet endroit est supérieure au seuil  $s^k(P)$ . La difficulté principale de cette méthode de détection du mouvement réside dans une construction robuste de chaque image  $R^k$  de la séquence de référence  $R$ .

### 3. construction de l'image de référence

Dans cette section, nous abordons la méthode de construction de l'image de référence par une définition des critères de réactualisation. Nous présentons ensuite le processus permettant sa mise à jour par l'exploitation d'une mesure permettant de juger de la stabilité temporelle de la référence.

#### 3.1. définition des critères de réactualisation

Dans le but d'avoir une image de référence qui réponde au mieux aux objectifs que nous nous sommes fixés, il nous faut définir des critères pertinents concernant son processus de réactualisation. Le mouvement des objets dans la scène provoque des perturbations dans la construction de l'image de référence. De plus, les objets peuvent s'arrêter dans la scène et intégrer la référence au bout

d'un certain temps. La figure 1 nous permet d'illustrer trois phénomènes différents. Elle comporte trois diagrammes binaires qui représentent respectivement le mouvement d'un objet, le changement dans l'état de la référence et la détection de l'objet en fonction du rang  $k$  dans la séquence. Ceux-ci sont numérotés de 1 à 3.

Cas 1 : Le premier phénomène concerne un objet qui traverse la scène sans s'arrêter (Fig. 1, entre les instants 0 et  $K_1$ ). L'objet doit être détecté pendant toute la durée du mouvement et, par conséquent, l'image de référence ne doit pas être modifiée durant cette période.

Cas 2 : Le deuxième phénomène concerne un objet en mouvement qui s'arrête dans la scène pendant un temps relativement court avant de reprendre son chemin (Fig. 1, entre les instants  $K_1$  et  $K_2$ ). L'objet doit être détecté pendant le mouvement, mais également durant le bref intervalle de temps durant lequel il s'est immobilisé. Il importe donc de ne pas modifier l'image de référence, même pendant l'arrêt momentané du mouvement.

Cas 3 : Le troisième phénomène correspond à un objet qui s'arrête définitivement dans la scène, ou pendant un temps suffisamment long pour qu'il soit nécessaire de l'oublier en tant qu'objet mobile et de l'intégrer dans le fond statique, donc dans l'image de référence (cf. Fig. 1, entre les instants  $K_2$  et  $K_3$ ). L'objet est en mouvement de  $K_2$  à  $K_{21}$  puis de  $K_{24}$  à  $K_3$ . La référence est mise à jour entre les instants  $K_{22}$  et  $K_{23}$  en y intégrant les contours de l'objet immobile; ceci conduit ainsi à une non détection de cet objet tant qu'il n'a pas repris son mouvement entre les instants  $K_{23}$  et  $K_{24}$ .

Ce dernier phénomène nous amène à définir le temps d'intégration d'un objet dans la référence, c'est-à-dire le temps nécessaire pour qu'un objet qui s'arrête dans la scène soit incorporé à l'image de référence (Fig. 1, entre les instants  $K_{21}$  et  $K_{23}$ ). Durant ce temps, la stratégie de réactualisation de l'image de référence peut revêtir de multiples facettes. Le choix de cette stratégie peut néanmoins se faire en fonction des deux objectifs suivants.

Le rôle premier de notre outil est la détection instantanée d'un objet entrant dans le champ de la caméra. Cette détection doit être maintenue pendant un certain temps, même si l'objet s'arrête. Pendant ce temps, l'image de référence ne doit donc pas évoluer. La partie gauche de la figure 2 illustre ce premier objectif

(Partie a). La courbe a1 de cette figure illustre l'évolution de la valeur de l'image courante des contours  $G^k$  au pixel  $P$  en fonction des instants d'acquisition, en supposant que le pixel  $P$  est choisi de telle façon que l'objet qui s'arrête dans la scène crée un contour à cet endroit. Ce cas constitue évidemment un cas idéal théorique. A l'instant  $k$ , l'amplitude  $G^k(P)$  du gradient au pixel  $P$  présente un front d'amplitude  $G_{Max}$ . La courbe a2 représente les valeurs que doit prendre l'image de référence des contours  $R^k$  au pixel  $P$  en fonction des instants d'acquisition. En effet, si l'on veut atteindre l'objectif 1, il ne faut réactualiser l'image de référence qu'après une période de  $K$  instants d'acquisition. Cette durée, que nous avons appelée « temps d'intégration », correspond à la durée maximale pendant laquelle un objet immobilisé momentanément dans la scène reste détecté. A l'instant  $(k + K)$ ,  $R^{k+K}(P)$  prend la valeur du gradient courant à l'instant  $k$  et l'objet n'est plus détecté. L'équation de réactualisation qui permet de satisfaire cette contrainte peut s'écrire :

$$R^k(P) = G^{k-K}(P). \quad (3)$$

Cette équation fait penser à un retard pur de valeur  $K$ . Ce retard consiste, d'un point de vue pratique, à stocker  $K$  images en mémoire, ce qui est très contraignant, pour des valeurs élevées de  $K$ , au niveau des capacités mémoire nécessaires.

En réalité, comme nous l'avons indiqué précédemment, la courbe a1 de la figure 2 n'a pas vraiment l'allure d'un échelon parfait, mais ressemble plutôt à un saut d'amplitude bruité. Ceci est fortement gênant puisque la valeur de  $G^k(P)$  peut être différente de  $G_{Max}$  et donc introduire une erreur dans la mise à jour de l'image de référence au pixel  $P$ . Il convient donc de baser la réactualisation de l'image de référence des contours sur plusieurs images de la séquence  $G$ . Ceci permet ainsi de filtrer le bruit perturbant l'apparition des contours. Le deuxième objectif à atteindre est, par conséquent, le filtrage du bruit qui entache l'apparition d'un contour au pixel  $P$ . Il ne faut plus baser la mise à jour de l'image de référence sur une seule image, mais sur un nombre suffisant d'images pour réaliser un filtrage temporel de ce bruit. On suppose désormais que le système de réactualisation de l'image de référence se comporte comme un filtre. L'entrée du filtre relative au pixel  $P$  est constituée des valeurs prises par les images de la séquence  $G$  au pixel  $P$ , tandis que la sortie représente les valeurs prises par les images de la séquence  $R$ . Pour atteindre l'objectif 2, il faut choisir un filtre qui lisse le signal sur une période donnée. Nous écartons la moyenne glissante car elle impose par nature une mémorisation des images. Nous choisissons un filtre qui réalise une intégration, tel un filtre du 1<sup>er</sup> ordre dont la réponse à un échelon de gradient d'amplitude  $G_{Max}$  peut s'écrire [2] :

$$R^{k+1}(P) = f.R^k(P) + (1 - f).G_{Max}. \quad (4)$$

$R^{k+1}(P)$  est la valeur mise à jour de  $R^k(P)$  une fois que la  $k^{ième}$  image de la séquence a été prise en compte. Le coefficient  $f$  permet d'ajuster le temps de réponse du filtre ( $f \in [0, 1]$ ). La partie droite de la figure 2 illustre la validation de ce second objectif de deux façons choisies parmi d'autres approches (Partie b). La courbe b1 représente l'évolution des valeurs prises par l'image courante des contours  $G^k$ . La courbe b2 indique les valeurs prises par l'image de référence des contours  $R^k$  au pixel  $P$  pour un filtre du premier ordre. La courbe b3 représente les valeurs de  $R^k(P)$  obtenues

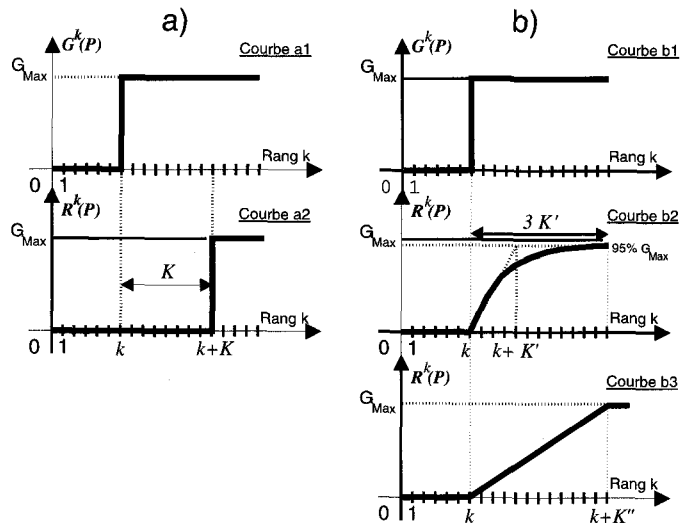


Figure 2. – Objectifs : a) Détection, b) Filtrage du bruit.

en utilisant un filtre moyenneur de largeur  $K''$ . A l'instant  $k$ ,  $G^k(P)$  représente toujours un front d'amplitude  $G_{Max}$ . La valeur de l'image de référence au pixel  $P$  tend vers la valeur finale  $G_{Max}$  en suivant l'équation de récurrence (4). Des propriétés simples des filtres du premier ordre, on déduit que la valeur prise par l'image de référence des contours au pixel  $P$  à l'instant  $(k + 3.K')$  est de l'ordre de 95% de  $G_{Max}$  où  $K'$  est la constante de temps du filtre qui dépend de  $f$  [13]. Pour la courbe b3, le temps  $K''$  pour que la valeur de l'image de référence tende vers la valeur finale dépend du nombre d'images considérées pour calculer la moyenne.

Pour choisir un filtre efficace dont le temps de calcul soit compatible avec l'objectif temps réel, il faut satisfaire simultanément les deux objectifs que nous venons de définir. Nous proposons un filtre récursif qui intègre la notion de retard pur et le filtrage du bruit. Nous détaillons le principe de ce filtre au paragraphe suivant.

### 3.2. processus de mise à jour

Chaque image  $R^k$ , représentant les contours statiques contenus dans la scène au rang  $k$ , est construite de manière récursive en utilisant le principe de la mise à jour présenté à la section précédente (cf. équation (4)). Notre contribution consiste à modifier le coefficient  $f$  au fur et à mesure du déroulement de la séquence, et ceci de façon locale pour chaque pixel  $P$ . L'équation de récurrence (4) devient alors :

$$R^{k+1}(P) = f^k(P).R^k(P) + [1 - f^k(P)].G^k(P). \quad (5)$$

Dans cette expression,  $f^k(P)$  représente une mesure de stabilité qui prend ses valeurs, pour chaque pixel de l'image, dans l'intervalle  $[0, 1]$ . L'évolution des valeurs de ce coefficient de stabilité est contrôlée par les variations de la différence entre l'image

de référence des contours et l'image courante des contours. Ce coefficient donne une mesure de la stabilité temporelle de l'image de référence à cet endroit.

Le premier objectif, que nous avons défini à la section précédente, peut être atteint en utilisant un retard pur au niveau de la réponse du système. En ajustant de façon judicieuse l'évolution de la mesure de stabilité  $f^k(P)$ , on peut de manière artificielle satisfaire facilement ce critère. Cette particularité s'avère très intéressante d'un point de vue pratique car elle simule un retard pur sans devoir stocker des images en mémoire. La figure 3 illustre un exemple d'évolution du coefficient  $f^k(P)$  selon un procédé du premier ordre à grande constante de temps. Cette stratégie permet de s'approcher du retard pur au niveau de la réponse de l'image de référence au pixel  $P$ . Nous constatons, sur la courbe d'évolution de la référence à l'instant  $k$ , que la pente à l'origine est proche de zéro.

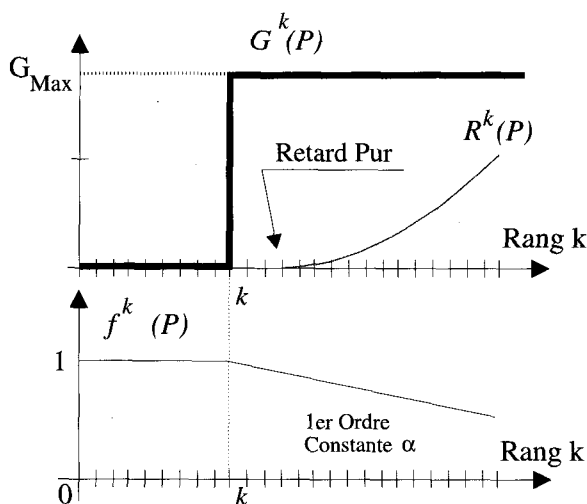


Figure 3. – Evolution de la référence et du coefficient de stabilité.

### 3.3. calcul du coefficient de stabilité

Le coefficient de stabilité  $f^k(P)$  au pixel  $P$  de l'image de rang  $k$  dépend de l'évolution des valeurs de l'image des variations de gradient  $D^k$  en ce pixel. Les valeurs de l'image  $D^k$  (cf. équation (1)) sont utilisées pour calculer l'image de détection des points de contours en mouvement  $C^k$  (cf. équation (2)). Le rôle du coefficient que nous allons introduire est de fournir à chaque instant une information sur la stabilité de l'image de référence en cours de construction. Il s'apparente à une notion de degré de confiance auquel peut prétendre un point de l'image de référence à un instant donné. Nous choisissons de faire varier le coefficient  $f$  selon des fonctions du premier ordre afin de contrôler facilement les temps de réponse du système. Deux expressions de  $f$  sont nécessaires pour différencier d'une part une augmentation du coefficient en cas de stabilité de la référence et, d'autre part, une diminution de ce coefficient dans le cas contraire. On notera  $\alpha$  et  $\beta$  les constantes de temps des deux fonctions du premier ordre.

S'ils sont choisis dans la plage  $[0, 1]$ , les variations du coefficient de stabilité sont limitées à l'intervalle  $[0, 1]$ . Une fonction  $\delta^k(P)$  représentant le détecteur temporel de mouvement est utilisée pour la mise à jour de l'image de référence. Son but est la détection du passage d'un objet au pixel  $P$ , comme nous l'expliquons au paragraphe suivant. L'actualisation du coefficient  $f$  suivant la valeur binaire de la fonction  $\delta^k(P)$ , peut alors s'écrire :

$$f^{k+1}(P) = \delta^k(P) \cdot [\alpha \cdot f^k(P)] + (1 - \delta^k(P)) \cdot [(1 - \beta) + \beta \cdot f^k(P)] \quad (6)$$

avec :

$$\delta^k(P) = \begin{cases} 1 & \text{si } D^k(P) > T \\ 0 & \text{sinon.} \end{cases} \quad (7)$$

$f^{k+1}(P)$  est la valeur mise à jour de  $f^k(P)$  une fois que la  $k^{\text{ième}}$  image de la séquence a été prise en compte.  $f^k(P)$ , coefficient de stabilité local, est ainsi mis à jour pour chaque image  $k$  de la séquence.

Le choix de la fonction  $\delta^k(P)$ , c'est-à-dire du seuil  $T$ , est important puisqu'il conditionne l'évolution du coefficient  $f^k(P)$  et par conséquent la robustesse de l'image de référence au pixel  $P$ . Nous cherchons à obtenir une information binaire qui reflète la présence ou l'absence de mouvement en chaque pixel  $P$ . Le seuil global  $T$  permet de différencier les pixels du fond des pixels affectés par le mouvement. Pour choisir le seuil  $T$ , nous procédons à une série de mesures visant à quantifier ces bruits. On suppose que le niveau de gris au pixel  $P$  dans l'image gradient  $G^k$  est la somme d'un niveau de gris déterministe noyé dans un bruit blanc. En supposant que le bruit est une variable aléatoire normale de moyenne nulle et de variance  $\sigma^2$ , la probabilité pour que le bruit soit compris entre  $-3\sigma$  et  $3\sigma$  est de l'ordre de 0,99. Il convient de choisir le seuil  $T$  de manière à tolérer la plupart des bruits. Pour déterminer l'écart-type  $\sigma$ , on choisit quelques points considérés comme fixes dans l'image, c'est-à-dire qui correspondent à des zones non affectées par le mouvement, telles que les immeubles, terre-plein, ... Le seuil  $T$  est choisi égal au triple de l'écart type maximum pour plusieurs séquences. En utilisant cette procédure, il est apparu qu'une valeur de seuil  $T$  égale à 10 niveaux de gris sur 256 pouvait être choisie pour l'ensemble des images recueillies dans le cadre de cette étude.

### 3.4. réglage des paramètres du modèle

Le réglage des paramètres de la procédure de réactualisation de l'image de référence est un problème délicat qu'il convient d'aborder. De manière à contrôler le temps nécessaire à l'intégration d'un objet dans la référence et ainsi permettre le réglage des paramètres  $\alpha$  et  $\beta$ , nous simulons l'apparition d'un objet sur un fond composé de deux parties, puis l'arrêt de cet objet dans la scène. Pour réaliser ce test, nous avons utilisé une séquence d'images synthétiques représentant un disque en déplacement dans la scène.

Nous étudions le comportement d'un pixel  $P$  qui se trouve sur un des points du contour du disque lorsque celui-ci s'arrête dans la scène. La valeur de l'image courante des contours  $G^k$  au pixel  $P$  va être modifiée. En effet, lorsque le disque s'arrête dans la scène, le signal observé dans l'image  $G^k$  au pixel  $P$  prend la forme d'un échelon d'amplitude  $G_{Max}$ . La valeur de l'image de référence des contours  $R^k$  au pixel  $P$  tend vers la valeur courante

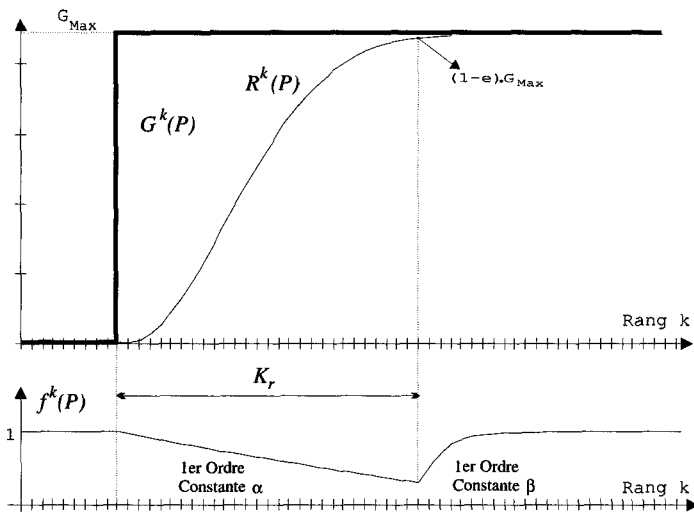


Figure 4. – Réponse du modèle réel à un échelon d'amplitude du gradient.

du gradient après un temps de réponse qu'il nous est possible de mesurer. Considérons le temps de réponse à  $e\%$  du modèle, que nous noterons  $K_r$ , mesuré en périodes d'acquisition. La figure 4 illustre l'évolution des valeurs prises par  $G^k(P)$  et  $R^k(P)$  ainsi que l'évolution du coefficient de stabilité  $f^k(P)$  dont la décroissance est exponentielle en fonction de l'instant d'acquisition  $k$ .

A l'instant initial ( $k = 0$ ), on suppose que la mesure de stabilité  $f^0(P)$  est stabilisé à la valeur 1, héritant ainsi d'un degré de confiance maximal. Avec  $\delta^k(P) = 1$ , l'équation de récurrence (6) peut alors s'écrire :

$$f^k(P) = (\alpha)^k. \tag{8}$$

Après l'acquisition de  $K_r$  images, nous voulons que la valeur de l'image de référence des contours  $R^k$  au pixel  $P$  soit inférieure ou égale à un certain pourcentage de  $G_{Max}$ . On notera  $e$  ce pourcentage qui correspond au rapport entre la valeur de consigne  $G_{Max}$  et la différence entre la valeur atteinte après  $K_r$  instants d'acquisition et la valeur de consigne. Compte tenu de cette définition de l'erreur, l'équation (5) devient alors :

$$G_{Max} \cdot (1 - e) \geq f^{K_r}(P) \cdot R^{K_r}(P) + [1 - f^{K_r}(P)] \cdot G_{Max}. \tag{9}$$

On a donc, par simplification, et en remplaçant  $f^k(P)$  par la valeur trouvée à l'équation (8) :

$$\begin{aligned} e &\leq -f^{K_r}(P)[R^{K_r}(P) - G_{Max}] \\ &\leq -(\alpha^{K_r})[R^{K_r}(P) - G_{Max}] \\ &\leq -(\alpha^{K_r})(\alpha^{K_r-1}) \dots (\alpha^{K_r-K_r})[R^0(P) - G_{Max}] \end{aligned}$$

d'où finalement :

$$e \leq \prod_{i=0}^{K_r} (\alpha^i). \tag{10}$$

Cette équation nous permet de calculer la valeur à donner au paramètre  $\alpha$  de façon à imposer le temps de réponse du modèle. La valeur de  $\alpha$  peut se mettre sous la forme :

$$\alpha = \exp \frac{2 \cdot \ln e}{K_r \cdot (K_r + 1)} \tag{11}$$

Pour choisir  $\alpha$ , il suffit donc de fixer le temps de réponse du système, c'est-à-dire le temps d'intégration d'un objet dans la référence ainsi que l'erreur finale admise. Le temps d'intégration est évidemment déduit du temps de calcul du détecteur; il est donné en périodes d'acquisition ( $K_r$ ).

Le paramètre  $\beta$  doit être relativement faible de manière à faire croître la mesure de stabilité rapidement lorsque l'erreur entre le gradient de référence et le gradient courant devient inférieure au seuil  $T$ . On choisira donc le paramètre  $\beta$  très petit devant  $\alpha$ . Son rôle est d'assurer une continuité dans l'évolution de la mesure de stabilité tout en filtrant les petites fluctuations du signal gradient dues aux bruits.

## 4. calcul du seuil local de détection

Le coefficient  $f^k(P)$  permet d'ajuster de manière dynamique le seuil  $s^k(P)$  utilisé pour le calcul de l'image de détection de mouvement (cf. équation (2)). Ce seuillage local de l'image des variations de gradient  $D^k$  permet d'extraire les points de contours affectés par le mouvement dans la scène.

L'approche envisagée consiste à utiliser la valeur du coefficient  $f^k(P)$  pour ajuster le seuil  $s^k(P)$  en exploitant au maximum la notion de stabilité temporelle que nous avons associée à ce coefficient. Le seuil local de détection  $s^k(P)$  est calculé de la façon suivante :

$$s^k(P) = s_2 - s_1 \cdot f^k(P). \tag{12}$$

Les valeurs  $s_1$  et  $s_2$  sont constantes.  $s_1$  représente la plage totale des variations possibles de  $s^k(P)$  tandis que  $s_2$  est la valeur maximale que ce seuil peut atteindre. Ces deux paramètres sont ajustés en fonction des conditions d'éclairage de la scène. En effet, une bonne stabilité temporelle de la référence, c'est-à-dire un coefficient de stabilité proche de 1, conduit à un seuil de détection plus bas et proche de ( $s_2 - s_1$ ). Au contraire, un coefficient de stabilité proche de 0 nous amène à être prudent quant à la décision à prendre et donc à augmenter le seuil de détection à une valeur proche de  $s_2$ .

Le choix des valeurs de  $s_1$  et  $s_2$  est important. Il faut que le seuil minimum de détection ( $s_2 - s_1$ ) soit légèrement plus élevé que 3 fois l'écart-type du bruit présent dans l'image, c'est-à-dire le seuil  $T$ . Enfin,  $s_1$  est la plage des variations possibles du seuil. Nous rappelons que la valeur de ce paramètre dépend du degré de prudence qu'il convient d'attribuer à la décision. Choisir une plage de variations trop grande revient à seuiller la valeur  $D^k(P)$  avec un seuil faible. Ceci peut être dangereux lorsque le coefficient  $f^k(P)$  est proche de zéro. Des tests sur des scènes d'extérieur nous ont conduits à fixer  $s_1$  à une valeur de 20 niveaux de gris (cf. section 5). Cette technique de seuillage permet en outre la prise en compte des variations locales d'éclairage. En effet, les changements locaux d'intensité lumineuse entraînent de faibles variations de la valeur des gradients. Ces variations perturbent l'image de référence et par conséquent contribuent à faire décroître le coefficient de stabilité. Le seuil de détection tend alors à augmenter et permet ainsi d'éviter de détecter les effets de ces variations locales de l'éclairage.

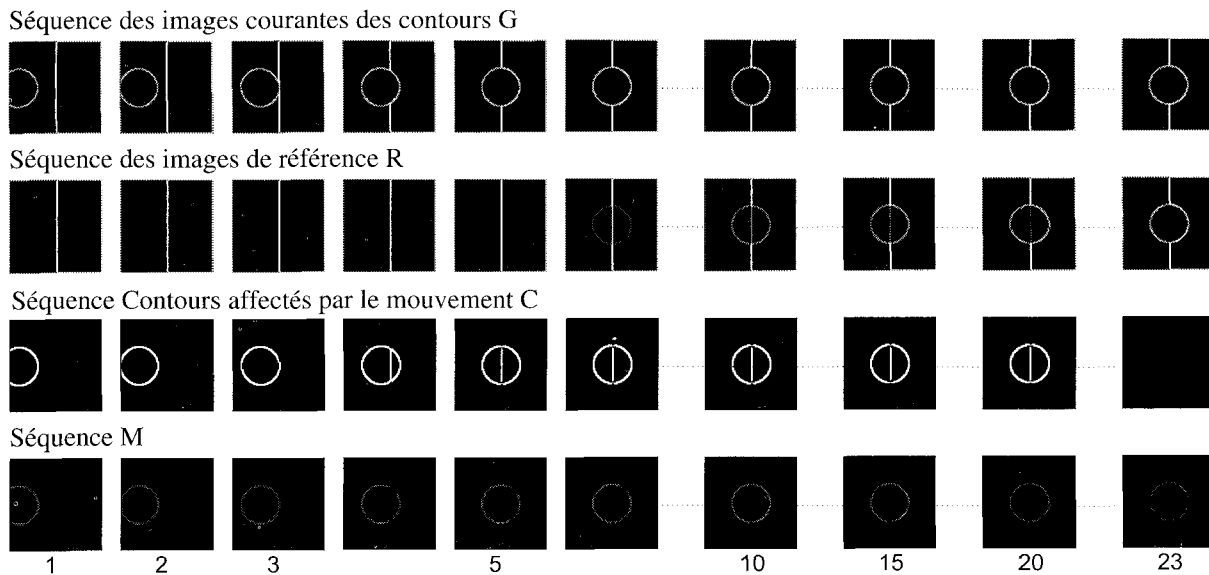


Figure 5. – Construction de l'image de référence.

## 5. résultats

### 5.1. simulation sur une séquence synthétique

A partir de la séquence synthétique précédemment définie, on se propose d'étudier le comportement de l'image de référence en fonction de l'instant d'acquisition  $k$  lorsqu'un objet s'arrête dans la scène. Pour faire cet essai, le paramètre  $\alpha$  est réglé à 0.98 et l'erreur  $e$  à 5%. La figure 5 illustre l'évolution des trois séquences respectives  $G$ ,  $R$  et  $C$ . Lorsque le disque s'arrête (image de rang  $k = 5$ ), l'image de référence est modifiée de manière à intégrer ses contours. Ceux-ci font partie intégrante de l'image de référence à l'instant d'acquisition  $k = 23$ , entraînant une non détection du disque dans l'image  $C^{23}$ . Le temps d'intégration correspond ici à 18 images (23-5), temps qui est ajusté par les réglages de  $\alpha$  et  $e$  (cf. équation (11)). Le paramètre  $\beta$ , réglé dans notre cas à 0.5, n'intervient pas dans le calcul de ce temps d'intégration. Les valeurs des paramètres  $\alpha$ ,  $\beta$  et  $s_1$  seront réglées respectivement à 0.98, 0.5 et 20 pour toutes les séquences à traiter. Sur cette séquence, on remarque que l'algorithme détecte les contours du disque en mouvement lorsque celui-ci se déplace sur un fond uniforme, c'est-à-dire entre les images 1 et 3. Pour les images 4 à 20, on détecte les contours en mouvement et les contours des zones affectées par le mouvement. Il est possible de s'affranchir simplement des contours statiques du fond en calculant une nouvelle image qui résulte de la multiplication pixel à pixel entre l'image  $D^k$  et l'image courante des contours  $G^k$ . Notons  $M^k$  cette nouvelle image ainsi calculée. La séquence composée des  $N$  images  $M^k$ , est présentée sur la figure 5. Sur cette séquence, on constate que les contours statiques du fond de la scène n'apparaissent plus.

### 5.2. expérimentation avec des séquences réelles

Nous présentons quelques résultats obtenus par la méthode de détection de mouvement décrite aux paragraphes précédents sur deux types de séquences dans le cas d'images urbaines. Ces séquences permettent de juger des performances de la procédure de seuillage adaptatif.

#### 5.2.1. séquence Urbaine 1

Une première séquence d'images urbaines est proposée de manière à analyser le principe de base de l'algorithme. Le traitement de ces images, dont certaines sont représentées à la figure 6, permet de mettre en évidence les différentes étapes de construction de l'image de détection des points de contours affectés par le mouvement. Le choix de cette séquence est motivé par la prise de vue qui permet la surveillance d'un carrefour dans sa globalité. La caméra est située sur un portique destiné à supporter les feux tricolores. Les images 1 et 2 de la séquence montrent deux véhicules immobilisés au feu rouge. En haut à droite de l'image, on aperçoit le flux de véhicules circulant sur l'autre partie du carrefour. Les images de la séquence  $G$  sont obtenues en appliquant un opérateur de Sobel sur les images de la séquence  $I$ . L'ensemble des images de la séquence  $G$  est utilisé pour construire l'image de référence des contours  $R^k$ . On remarque que l'image de référence ne contient pas les contours des véhicules immobilisés pendant un temps relativement court. Les images de variations de gradient  $D^k$  sont obtenues en réalisant la différence entre les images de la séquence  $G$  et l'image  $R^k$ . La figure 6 illustre également les résultats obtenus pour cette même séquence lors de la traversée du carrefour par un piéton. Nous avons choisi deux images originales prises à des instants différents dans la séquence (images



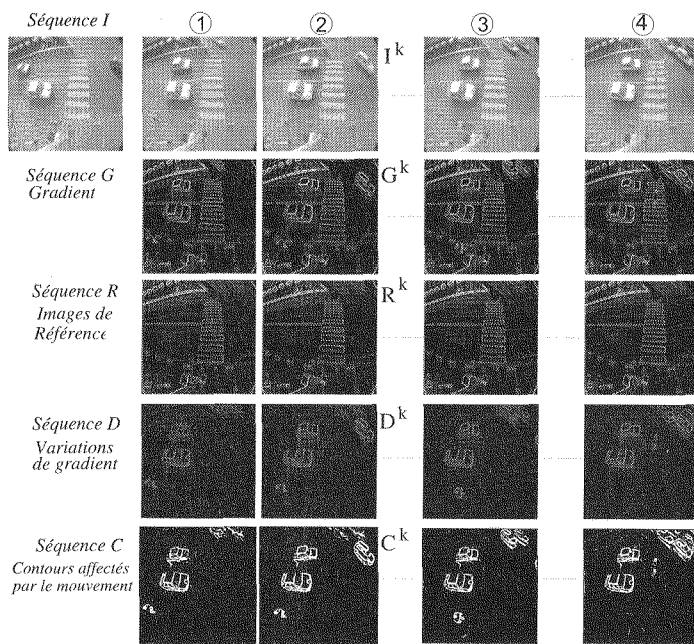


Figure 6. – Séquence Urbaine 1.

3 et 4). L'extraction des points de contours affectés par le mouvement est ensuite obtenue par un seuillage à une valeur de  $s_2$  de 30 niveaux de gris. Pour cette valeur du seuil, on remarque que les ombres portées par les véhicules sur la chaussée, visibles sur les images originales et les images de variations de gradient, ne sont pas détectées. On constate cependant la présence de bruit, c'est-à-dire de points de contours qui n'auraient pas dû être détectés. Ces points isolés peuvent être néanmoins supprimés par une procédure de filtrage morphologique.

### 5.2.2. séquence Urbaine 2

La figure 7 illustre les résultats obtenus pour la séquence urbaine 2. Cette séquence correspond à la traversée d'un carrefour par un groupe de piétons. On remarque facilement que l'ensemble des piétons est correctement détecté. Sur cette figure, les images originales de la séquence  $I$  ainsi que les images des variations de gradient de la séquence  $D$  sont proposées. Le seuillage adaptatif local proposé, pour des valeurs respectives de  $s_1$  et  $s_2$  de 20 et 30 niveaux, permet de mettre en évidence les objets en mouvement dans la séquence. Sur ce carrefour, le flux important des véhicules et des piétons entraîne une instabilité de l'image de référence. Ceci peut malheureusement conduire à quelques fausses détections.

### 5.3. robustesse vis-à-vis des variations d'éclairage

L'algorithme proposé a été utilisé dans le cadre du programme de recherche et développement PRAXITELE pour lequel il s'agissait de réaliser une maquette opérationnelle d'un système de détection

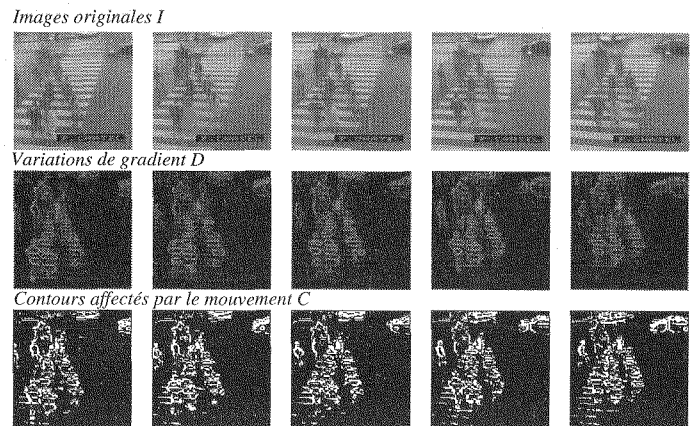


Figure 7. – Séquence Urbaine 2.

d'intrusions sur un parking par vidéo-surveillance. Le parking, dédié au stationnement de véhicules électriques en location, est doté de bornes de courant qui permettent la recharge des batteries des véhicules. Pour faciliter la location des véhicules, une station est installée à l'entrée du parc permettant ainsi l'identification de l'utilisateur. Le détecteur de mouvement proposé est utilisé de manière à gérer le contrôle des accès au parking [15]. Lors des différentes procédures mises en place pour l'évaluation du prototype de vidéo-surveillance, les problèmes rencontrés sont de plusieurs types. Les principales difficultés proviennent des ombres portées par les objets dans la zone de surveillance, des reflets de sources lumineuses (soleil ou projecteur) sur la carrosserie des véhicules et des reflets des nuages sur les pare-brise des véhicules. Deux sites de tests ont permis l'évaluation du prototype dans des conditions d'éclairage et météorologiques fortement variables. La figure 8 présente quelques images de détection d'intrusions enregistrées lors de tests de jour, comme de nuit. Elle comporte les images originales sans intrusion ainsi que les images de détection des contours des objets en mouvement lors d'intrusions de véhicules ou de piétons. Les réglages des constantes de temps  $\alpha$  et  $\beta$  n'influent que très peu sur les résultats dans le cadre de cette application. Par contre, la valeur du seuil  $s_2$  a été augmentée

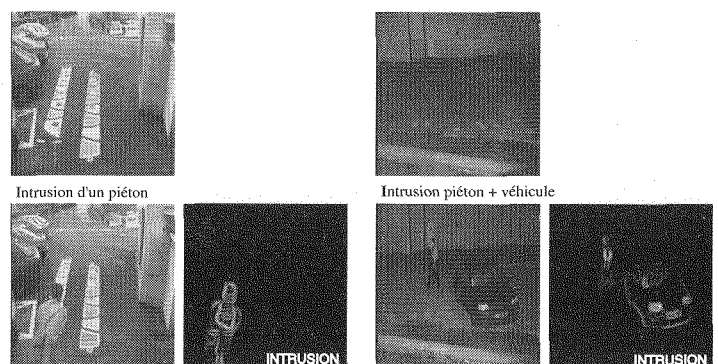


Figure 8. – Tests de détection d'intrusions de jour (partie gauche) et de nuit (partie droite).



à 40 niveaux pour pouvoir diminuer au maximum le nombre de fausses détections.

Le tableau 1 présente les résultats obtenus lors d'une campagne de tests. Il indique le nombre réel d'intrusions dans la zone de surveillance, le nombre de fausses alarmes, le nombre de détections trouvées par le système proposé ainsi que le pourcentage de bonnes détections dans trois conditions de prise de vue (soleil, pluie et nuit).

On remarque que le pourcentage de bonnes détections est dans tous les cas supérieur à 96%. Cette évaluation a été réalisée sur des périodes de trois heures en dépouillant en temps différé les images enregistrées sur magnéto-scope.

Tableau 1. – Résultats des tests de détection pour PRAXITELE.

Conditions climatiques	Soleil	Pluie	Nuit
Nombre d'intrusions	127	33	6
Fausses Alarmes	0	1	0
Détections	123	32	6
Pourcentage de bonnes détections	96.8%	96.9%	100%

## 6. conclusion

Les travaux présentés dans cet article ont permis d'aborder la détection du mouvement dans une séquence d'images monoculaires dans un contexte où la caméra est stationnaire. Ils reposent sur une combinaison des informations spatiales et temporelles contenues dans la séquence. Celles-ci sont englobées dans une modélisation permettant de s'adapter au mieux aux différentes situations résultant de l'éclairage non uniforme d'une scène naturelle, de l'hétérogénéité des fonds d'une scène en milieu urbain et du mouvement non uniforme des objets. Le détecteur proposé exploite les différences spatiales et temporelles existant entre une image de référence construite et réactualisée en ligne et chaque image de la séquence étudiée. La construction et la mise à jour de l'image de référence représentant les contours statiques de la scène a été menée en respectant les contraintes suivantes :

- un retard dans la mise à jour de cette image,
- un contrôle du temps d'intégration des objets dans la référence,
- un filtrage du bruit créé par l'apparition des contours.

Ces contraintes ont été respectées grâce au calcul d'un coefficient de stabilité tout en évitant un stockage important d'images. La stabilité de l'image de référence, estimée en continu et de façon locale, permet également d'avoir un seuil de détection plus précis qu'un seuil global. Le filtre proposé possède les avantages d'un filtrage récursif dont la réponse est entièrement contrôlée quel que soit le type d'objets en mouvement. Les objets en mouvement dans la scène prennent l'apparence d'une image binaire représentant la présence ou l'absence de mouvement. Les résultats sur des séquences réelles acquises en extérieur ont montré la robustesse de l'algorithme proposé par rapport aux variations aussi bien globales que locales (ombres, phares des véhicules) des conditions d'éclairage de la scène.

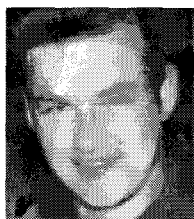
## BIBLIOGRAPHIE

- [1] Makarov A., Comparison of Background Extraction Based Intrusion Detection Algorithms, *International Conference on Image Processing, ICIP*, pages 521-524, Lausanne, Switzerland, September 16-19 1996.
- [2] Makarov A. and Vesin J.M. and Reymond F., Intrusion detection robust to slow and abrupt lighting changes. In *Real-Time Imaging*, volume 2661, pages 44-54, January 1996.
- [3] Rourke A. and Bell M.G.H., An image-processing system for pedestrian data-collection. In *7th International Conference on Road Traffic Monitoring and Control*, pages 123-126, London, April 26-28 1994.
- [4] Motamed C., *Application de la vision artificielle à la sécurité en robotique*, PhD thesis, Université de Technologie de Compiègne, Novembre 1992.
- [5] Vieren C., *Segmentation de scènes dynamiques en temps-réel. Application au traitement de séquences d'images pour la surveillance de carrefours routiers*, PhD thesis, Université des Sciences et Techniques de Lille, Avril 1988.
- [6] Bartolini F. and Cappellini V. and Giani C., Intelligent Image Processing Techniques for Urban Traffic Monitoring, In *IEEE book IEEE International Conference on IMACS, Computational Engineering in Systems Applications*, volume 4, pages 771-776, Lille, France, July 9-12 1996.
- [7] Cabestaing F., *Détection de contours en mouvement dans une séquence d'images. Conception et réalisation d'un processeur cablé Temps-Réel*. Ph D Thesis, Université des Sciences et Techniques de Lille, Janvier 1992.
- [8] Reading I.A.D. and Wan C.L. and Dickinson K.W., Developments in pedestrian detection, *Traffic Engineering and Control*, pages 538-542, October 1995.
- [9] Blosseville J.-M. and Bouzar S. and Lenoir F. and Glachet R., Traitement d'image pour la mesure et la surveillance du trafic : une avancée significative, *Revue TEC, Spécial télématique routière*, 126 : 29-34, 1994.
- [10] Karmann K.P. and Von Brandt A. and Gerl R., Moving object segmentation based on adaptative reference images, *Signal Processing*, 5 : 951-954, 1990.
- [11] Inigo R. M., Traffic Monitoring and Control Using Machine Vision : A Survey, *IEEE Transactions on Industrial Electronics*, IE-32 (3) : 177-185, August 1985.
- [12] Munkelt O. and Kirchner H. Stabil : A system for monitoring persons in image sequences. In *Image and Video Processing IV*, volume 2666, pages 163-179, February 1996.
- [13] Guyenot P. and Hans T., *Régulation et Asservissement*, Eyrolles, Paris, 180 p, 1989.
- [14] Vannoorenberghe P., *Détection de mouvement par analyse de séquences d'images monoculaires. Application à l'estimation de flux de piétons en milieu urbain.*, Université du Littoral, Juillet 1997.
- [15] Vannoorenberghe P. and Motamed C. and Blosseville J.-M. and Postaire J.-G., Automated video surveillance for access control. In CD-ROM, editor, *Intelligent Transportation System, Realizing the future*, Orlando, Florida, October 14-18, 1996.
- [16] Vannoorenberghe P. and Motamed C. and Blosseville J.-M. and Postaire J.-G., Motion Detection for Non-Rigid Objects. Application to Pedestrians Monitoring in an Urban Environment. In *IEEE International Conference on IMACS, Computational Engineering in Systems Applications*, volume 4, pages 438-442, Lille, France, 9-12 July, 1996.
- [17] Velastin S.A. and Yin J.H. and Davies A.C. and Vicencio-Silva M.A. and Allsop R.E. and Penn A., Automated measurement of crowd density and motion using image processing, In *7th International Conference on Road Traffic Monitoring and Control*, pages 127-132, London, 26-28 April 1994.
- [18] Haynes S.M., Detection of moving edges, *Computer Graphics and Image Processing*, 21, : 35-43, 1983.

Manuscrit reçu le 24 février 1997.

### LES AUTEURS

#### Patrick VANNOORENBERGHE



Patrick Vannoorenberghé est né en 1969. Il est titulaire d'un D.E.A. Productique de l'Université des Sciences et Technologies de Lille en 1993. Il a soutenu une thèse de Doctorat en Productique : Automatique et Informatique Industrielle à l'Université du Littoral-Côte d'Opale en Juillet 1997. Son activité de recherche s'est d'abord développée autour du traitement de séquences d'images monoculaires avec des applications dans le domaine de la surveillance de carrefours urbains. Depuis 1998, il est Attaché Temporaire d'Enseignement et de Recherche au Laboratoire d'Analyse des Systèmes du Littoral et ses travaux s'orientent vers la perception multicapteurs et la fusion de données.

#### Cina MOTAMED



Cina Motamed est née en 1965 à Téhéran. Depuis 1988, il est titulaire d'une Maîtrise d'Electronique, Electrotechnique et Automatique de l'université de Caen. Il a obtenu, en 1992 le titre de Docteur de l'UTC (Université de Technologie de Compiègne) dans la spécialité Contrôle des Systèmes. Ses intérêts portent sur la vision par ordinateur et plus particulièrement sur les systèmes de vidéo-surveillance. Maître de Conférence depuis 1993, à l'IUT de Calais au sein de l'université du Littoral-Côte d'Opale, il travaille au Laboratoire d'Analyse des Systèmes du Littoral.

#### Jack-Gérard POSTAIRE



Jack-Gérard Postaire, Ingénieur (I.D.N, 1970) et Docteur ès Sciences (USTL, 1981) est professeur en Automatique à l'Institut Universitaire de Technologie de Lille depuis 1982. Auparavant, il a été Maître de Conférences à la Faculté des Sciences de Rabat, Maroc, de 1976 à 1982, et Professeur Associé à l'Université de Sherbrooke au Canada, de 1972 à 1974. Il est auteur ou co-auteur de plus de 190 conférences et 65 articles scientifiques. Il a dirigé ou co-dirigé plus de 40 thèses et autant de projets scientifiques, principalement dans le domaine de l'analyse des données, de la classification automatique et du traitement et de l'analyse des images numériques.